

Endogenous retroviruses: **Still active after all these years?**

Jonathan P. Stoye

Embedded in the genomes of all vertebrates are the proviral remnants of previous retroviral infections. Although the overwhelming majority has suffered inactivating mutations, current research suggests that members of one family of human retroelements may still be capable of movement.

Address: National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK.
E-mail: jstoye@nimr.mrc.ac.uk

Current Biology 2001, 11:R914–R916

0960-9822/01/\$ – see front matter

© 2001 Elsevier Science Ltd. All rights reserved.

Retrovirus replication has two unique and characteristic properties: reverse transcription of viral RNA into DNA, and integration of the newly synthesized DNA into the genome of the infected cell. The latter property provides a means for retroviruses to colonize the germ line, as the progeny of the infected germ cell will inherit the provirus formed as an endogenous retrovirus [1]. Provided that the provirus retains replicative potential, amplification of copy number can occur with time by reinfection of, or by retrotransposition within, germ cells. Over time, the integrated provirus will accumulate random mutations and eventually only inactive fossilized proviruses will remain. In addition, homologous recombination between the two proviral long terminal repeats (LTRs) can occur, leading to the formation of so-called solo LTRs.

The vast majority of endogenous retroviruses are inactive. Most species have a few proviruses with replicative potential, however, and of these, a handful are associated with spontaneous tumours. Humans have appeared to be one exception to this generalization, and it is widely believed that none of our complement of proviruses is replication competent. Very recently, however, a few tantalizing clues suggesting the presence of potentially active human endogenous retroviruses (HERVs) have been published. In particular, one study published recently in *Current Biology* [2] has demonstrated the presence of a very recently acquired, apparently intact HERV in a small fraction of the human population.

Endogenous retroviruses have been found in all vertebrate species examined to date and constitute a significant fraction of the genome. In general, the number of proviruses found in a given species is roughly proportional to the effort expended in searching for them. From studies in humans [3], pigs [4] and mice [1] we can infer that each species has a limited number of retrovirus groups (perhaps

up to thirty). Each group will have undergone amplification independently and full-length proviruses are present in copy numbers ranging from one to a thousand or more, with perhaps ten to a hundred times more solo LTRs. Members of a given group will show a relatively high degree of sequence conservation, whereas a much lower degree of homology is shown between different groups. Each provirus is randomly integrated in the genome and the likelihood that two independent integrations occur at the same position is essentially negligible. Each provirus can therefore be identified via its host flanking sequences. This makes it relatively trivial to design PCR assays that will distinguish the presence or absence of a given provirus, thereby allowing studies of proviral inheritance.

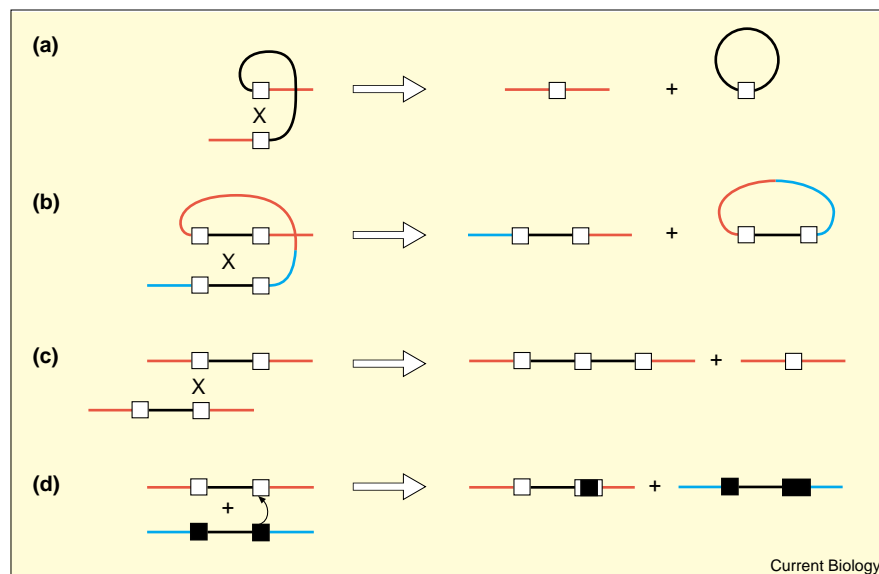
Proviral inheritance might have numerous consequences for the host. Some stem from the insertion of multiple copies of DNA sequences containing signals capable of modifying transcription or RNA processing. Thus proviruses might act to cause chromosomal rearrangement by homologous recombination, as a source of novel control sequences for cellular genes or as insertional mutagens. Alternatively, there might be consequences from viral gene expression, with either pathogenic [1] or possibly beneficial effects [5]. In the extreme case, transcription may lead to virus activation and the formation of virally induced tumours, as has been well documented with certain endogenous murine leukemia viruses and mouse mammary tumour viruses [1].

It was this latter possibility that initially fueled the search for HERVs with replicative potential. Later studies have been greatly facilitated by the progress of the human genome project. Collectively, HERVs comprise approximately 7% of the human genome [6,7]. Three classes of HERV have been identified: class I HERVs (2.3% of the genome) are similar to exogenous gamma-retroviruses such as murine leukemia virus; class II HERVs (0.7%) are related to beta-retroviruses such as mouse mammary tumour virus; and class III HERVs (4%) are distantly related to spuma-retroviruses. Despite their number, to date no replication-competent HERVs have been identified.

For a number of years the HERV-K(HML-2) group has looked the most likely to provide an active virus [8]. Currently, there are 30–50 relatively intact HERV-K(HML-2) proviruses, and around 1000 solo LTRs in the human genome. One or more proviruses belonging to this group are responsible for the production of virus-like particles from human teratocarcinoma cells. Initial analyses of the coding potential of cloned proviruses from this group, by sequencing or coupled *in vitro* transcription–translation

Figure 1

DNA recombination events involving endogenous retroviruses. (a) Recombination between the two LTRs of a single provirus resulting in the formation of a solo LTR. The remainder of the provirus will be lost. (b) Homologous recombination between two proviruses on the same chromosome resulting in a microdeletion and loss of the intervening sequences. (c) Recombination between the 3' and 5' LTRs of a given provirus leading to a tandemly duplicated provirus. (d) Gene conversion leading to non-homologous gene exchange with no proviral loss. Boxes indicate LTRs, coloured lines are cellular sequences.



Current Biology

assays, did not identify any intact proviruses. With one exception, all carried mutations interrupting one or more of the viral genes [9,10]. The one exception, a provirus mapped to human chromosome 7, carried a single point mutation in the *pol* gene, substituting cysteine for tyrosine in a highly conserved YXDD motif of the protein product [11]. The number of mutations is very low, however, contrasting significantly with the highly mutated state of almost all other HERVs, where multiple frameshifts litter the proviral genome [3]. This suggests that the HERV-K(HML-2) proviruses studied have entered the human genome relatively recently. This is consistent with the fact that many of these proviruses are not present in our closest primate neighbours [9,12].

This does not appear to be the end of the story, however. Very recent reports from two groups now indicate a degree of polymorphism in the human population, and have identified two apparently intact proviruses. One study [13] focussed on the chromosome 7 provirus and showed that two out of seven individuals carry proviruses without the inactivating tyrosine-to-cysteine mutation. The second study [2] identified two novel proviruses that are polymorphic in humans with allele frequencies of 4 and 19%. Characterization of these proviruses revealed that one had identical LTRs and complete open reading frames for all the viral proteins. Both these studies reveal one potential drawback of relying on a single sequenced allele as representative of the human genome.

The identity of the LTRs and the presence of open reading frames, coupled with the fact that these proviruses have not become fixed in the human population, suggest

that these proviruses might have formed within the last 200,000 years. If so, is it possible that reintegrations are continuing to occur, either in the germ line or in somatic tissues? One approach to this question is to look for the potential for movement — to test whether the proviruses identified can yield infectious virus following DNA transfection. One potential problem with this approach will be to find the right cells — a number of retroviruses will only grow on an extremely limited set of cells. In addition, inapparent mutations in non-coding sequences may be present.

Alternatively, one could carry out a detailed search for novel integrations in different human populations or tumour tissues. Unfortunately, the PCR approaches used in the experiments described above will not be suitable for such investigations since they require knowledge of the proviral flanking sequence. Rather, high-resolution Southern blotting approaches, using strategies designed to maximize the resolution of multicopy retroviral families [14] should be employed. Preliminary studies have confirmed the potential of this approach (J. Fitzpatrick and J. M. Coffin, personal communication). Such studies might yet suggest that endogenous retroviruses can provide a genetic basis for some human diseases.

Endogenous retroviruses show a significant tendency to recombine. Figure 1 illustrates a variety of different recombination products that have been documented. They include: homologous recombination between the viral LTRs, resulting in solo LTRs [15]; recombination between two different proviruses, leading to substantial deletions or rearrangements of cellular DNA [16] (J. Fitzpatrick and J.M. Coffin, personal communication); recombination

between the 3' and 5' LTRs of allelic proviruses, resulting in a tandemly duplicated provirus [13,17]; and gene conversion events which occur at a surprisingly high frequency [18]. There is, however, no mechanism for the precise excision of a provirus (except by gene conversion with another copy of the same region that never had a provirus).

A shared provirus in two species thus implies a common ancestor. One such provirus present in gorillas and chimpanzees, but not present in humans, provides evidence that some fraction of the former two species are more closely related to one another than to humans, even though the gorilla lineage diverged before separation of humans and chimps [19]. A second manifestation of the use of endogenous retroviruses in evolutionary studies is the use of sequence comparisons between the two LTRs of a provirus, identical at the time of integration, as molecular clocks. Phylogenetic analyses of tree structures involving both LTRs will readily reveal gene conversion events which will often escape detection in analyses involving single insert sequences, skewing estimates of mutation rates [18].

Although the HERV-K(HML-2) group still shows signs of movement, evolutionary studies show clearly that this group is of considerable antiquity. A number of proviruses are present at the same locations in humans and Old World monkeys, implying that these elements first entered the primate germ line over 35 million years ago [12,20]. How has the ability to generate novel integrants been preserved in the face of continued random mutagenesis? Conservative estimates suggest that the rate of mutation is around 0.2% per million years [18], a rate which would be expected to inactivate proviruses quite rapidly. Analyses of HERV-K(HML-2) LTRs formed at different times during evolution are consistent with the sequential appearance of a series of active proviruses, each seeding a novel burst of increase in provirus copy number [12,20].

How are these active proviruses formed? Given the high rates of gene conversion observed between HERVs it is tempting to speculate that recombination between proviruses at different chromosomal locations plays an important role in maintaining the endogenous life style, perhaps as important as recombination during reverse transcription for defective retroviruses [1]. The combination of random mutagenesis leading to alterations in proviral expression, together with gene conversion and/or recombination during reverse transcription to reverse inactivating mutations in coding sequences, might allow continued low level generation of novel proviruses capable of reintegration, thereby permitting survival of integrated retroviruses in the germ line.

Acknowledgements

I thank Michael Bock and John Coffin for helpful discussions.

References

- Boeke JD, Stoye JP: **Retrotransposons, endogenous retroviruses, and the evolution of retroelements.** In *Retroviruses*. Edited by Coffin JM, Hughes SH, Varmus HE. Cold Spring Harbor: Cold Spring Harbor Press; 1997: 343-435.
- Turner G, Barbulescu M, Su M, Jensen-Seaman MI, Kidd KK, Lenz J: **Insertional polymorphisms of full-length endogenous retroviruses in humans.** *Curr Biol* 2001, 11: 1531-1535.
- Tristem M: **Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the human genome mapping project database.** *J Virol* 2000, 74:3715-3730.
- Patience C, Switzer WM, Takeuchi Y, Griffiths DJ, Goward ME, Heneine W, Stoye JP, Weiss RA: **Multiple groups of novel retroviral genomes in pigs and related species.** *J Virol* 2001, 75:2771-2775.
- Mi S, Lee X, Li X-p, Veldman GM, Finnerty H, Racie L, LaVallie E, Tang X-Y, Edouard P, Howes S, et al.: **Syncytin, a captured retroviral envelope protein involved in human placental morphogenesis.** *Nature* 2000, 403:785-789.
- Smit AF: **Interspersed repeats and other mementos of transposable elements in mammalian genomes.** *Curr Opin Genet Dev* 1999, 9:657-663.
- Bock M, Stoye JP: **Endogenous retroviruses and the human germline.** *Curr Opin Genet Dev* 2000, 10:651-655.
- Löwer R, Löwer J, Kurth R: **The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences.** *Proc Natl Acad Sci USA* 1996, 93:5177-5184.
- Barbulescu M, Turner G, Seaman MI, Deinard AS, Kidd KK, Lenz J: **Many human endogenous retrovirus K (HERV-K) proviruses are unique to humans.** *Curr Biol* 1999, 9:861-868.
- Tönjes RR, Czauderna F, Kurth R: **Genome-wide screening, cloning, chromosomal assignment, and expression of full-length human endogenous retrovirus type K.** *J Virol* 1999, 73:9187-9195.
- Mayer J, Sauter M, Racz A, Scherer D, Mueller-Lantzsch N: **An almost-intact human endogenous retrovirus K on human chromosome 7.** *Nat Genet* 1999, 21:257-258.
- Medstrand P, Mager D: **Human-specific integrations of the HERV-K endogenous retrovirus family.** *J Virol* 1998, 72:9782-9787.
- Reus K, Meyer J, Sauter M, Scherer D, Müller-Lantzsch N, Meese E: **Genome organization of the human endogenous retrovirus HERV-K(HML-2.HOM) (ERV6) on chromosome 7.** *Genomics* 2001, 72:314-320.
- Frankel WN, Stoye JP, Taylor BA, Coffin JM: **A genetic linkage map of endogenous murine leukemia viruses.** *Genetics* 1990, 124:221-236.
- Seperack PK, Strobel MC, Corrow DJ, Jenkins NA, Copeland NG: **Somatic and germ-line reverse mutation rates of the retrovirus-induced dilute coat-color mutation of DBA mice.** *Proc Natl Acad Sci USA* 1988, 85:189-192.
- Sun C, Skaletsky H, Rozen S, Gromoll J, Nieschlag E, Oates R, Page DC: **Deletion of the azoospermia factor a (AZFa) region of human Y chromosome caused by recombination between HERV15 proviruses.** *Hum Mol Genet* 2000, 9:2391-2396.
- Pavelitz T, Liao D, Weiner AM: **Concerted evolution of the tandem array encoding primate U2 snRNA (the RNU2 locus) is accompanied by dramatic remodelling of the junctions with flanking chromosomal sequences.** *EMBO J* 1999, 18:3783-3792.
- Johnson WE, Coffin JM: **Constructing primate phylogenies from ancient retrovirus sequences.** *Proc Natl Acad Sci USA* 1999, 96:10254-10260.
- Barbulescu M, Turner G, Su M, Kim R, Jensen-Seaman MI, Deinard AS, Kidd KK, Lenz J: **A HERV-K provirus in chimpanzees, bonobos and gorillas, but not humans.** *Curr Biol* 2001, 11:779-783.
- Kurdyukov SG, Lebedev YB, Artamonova II, Gorodentseva TN, Batrak AV, Mamedov IZ, Azhikina TL, Legchilina SP, Efimenko IG, Gardiner K, Sverdlov ED: **Full-sized HERV-K (HML-2) human endogenous retroviral LTR sequences on human chromosome 21: map locations and evolutionary history.** *Gene* 2001, 273:51-61.